



# Artificial selection of microbial communities: what have we learnt and how can we improve?

Joshua L Thomas<sup>\*</sup>, Jamila Rowland-Chandler<sup>\*</sup> and Wenying Shou



Microbial communities are capable of performing diverse functions with important bioindustrial and medical applications. One approach to improving community function is to breed new communities by artificially selecting for those displaying high community function ('community selection'). Importantly, community selection can improve the function of interest without needing to understand how the function arises, just like in classical artificial selection of individuals. However, experimental studies of community selection have had varied and largely limited success. Here, we review a conceptual framework to help foster an understanding of community selection and its associated challenges, and provide broad insights for designing effective selection strategies.

## Address

Centre for Life's Origins and Evolution, Department of Genetics, Evolution and Environment, University College London, United Kingdom

Corresponding author: Shou, Wenying ([wenying.shou@gmail.com](mailto:wenying.shou@gmail.com))

<sup>\*</sup>Equal contribution.

Current Opinion in Microbiology 2024, 77:102400

This review comes from a themed issue on **Microbial Systems and Synthetic Biology**

Edited by **Victor Sourjik** and **Kiran Patel**

Available online 12 December 2023

<https://doi.org/10.1016/j.mib.2023.102400>

1369–5274/Crown Copyright © 2023 Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

## Introduction

Microbial communities are important for human health and the biogeochemical cycling of elements. Members of a community *interact* when one member affects the physiology of another by, for example, releasing nutrients or toxins [1–3]. Consequently, microbial communities can perform *community functions*, defined as community-level activities that are not equal to the sum of monoculture activities (Figure 1a). Examples include the bioindustrial synthesis of certain natural products [4,5], protection against pathogens [6], and the degradation of harmful

waste compounds [7,8]. Note that if a community-level activity is the sum of monoculture activities, then we can simply focus on individual species, as the community context is no longer necessary.

Community function can be improved using two broad approaches. In the 'bottom-up' approach, microbial communities can be constructed with a set of engineered species, each optimized to perform an activity contributing to the community function [9]. However, this requires a mechanistic understanding of interactions and how they give rise to community function, which is often unknown. Alternatively, in the 'top-down' approach, community function can be improved via artificial *community selection*: directed evolution of communities to achieve a high community function. Importantly, the top-down approach does not require an understanding of how interactions generate community functions, yet interactions that drive community function can be revealed by comparing evolved versus ancestral communities.

A community selection experiment consists of multiple selection cycles. Each cycle (Figure 1b) starts with a set of 'Newborn' communities (Newborns) at low cell densities, which are allowed to grow over a period of 'maturation' time defined by the experimentalist (olive arrows) into 'Adult' communities (Adults). During community maturation, cells interact, proliferate, and possibly mutate. At the end of a cycle, Adults displaying the highest community function are chosen to 'reproduce' (pink arrows): each is used to inoculate multiple Newborns for the next selection cycle.

In principle, community selection should improve community function. However, experimental community selection studies have shown varied and largely limited success ([10–13,16–19,20], reviewed in Refs. [21–24]). Often, community selection failed to improve community function compared with selection for low function or random selection, with some studies even reporting a decrease in community function despite selection. This begs the question: why might community selection have limited efficacy in experimental systems?

In this review, we first build a conceptual framework to facilitate an understanding of community selection and what makes it challenging. Second, we aim to provide broad insights to help experimentalists devise effective

community selection strategies. For conceptual clarity, we also provide a glossary of key terms.

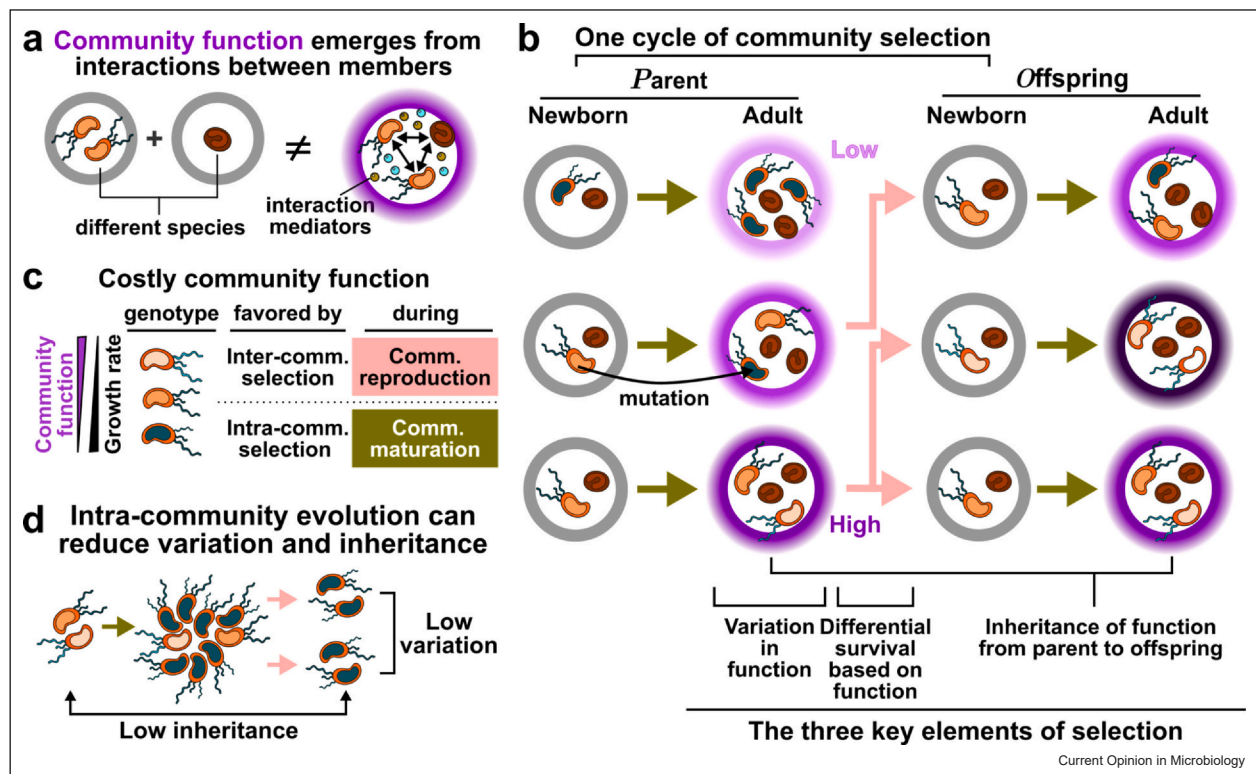
## The Jekyll-and-Hyde nature of community selection

Evolution can operate on any type of biological entity, that is, any structure with a boundary such that the birth, growth, survival, or death of one entity is separable from that of other similar entities [25]. Entities can be molecules, organelles, cells, or communities. Successfully selecting a trait has three requirements: variation in the trait among entities, differential survival of entities

based on the trait, and inheritance of the trait from parent to its offspring [26] (Figure 1b, bottom).

Despite what the name might suggest, community selection involves the selection on two types of entities: communities and individuals. Communities are the substrate for inter-community selection, which acts during community reproduction (pink arrows in Figure 1b) and favors high community function ('Dr. Jekyll'). In contrast, individual cells are the substrate for intra-community selection, which acts during community maturation (olive arrows in Figure 1b) and favors fast-growing individuals that may be deleterious to community function ('Mr. Hyde').

Figure 1



The artificial selection of microbial community function. **(a)** Community functions emerge from interactions among community members. Thus, the function of a community (purple halo) does not equal the sum of the activities of monocultures. For example, in a starch-degrading community [11], all species could produce starch-degrading enzymes, but the total activity of a community differed from the sum of monocultures due to interactions that affected species growth. As another (strict-sense) example, a community of *Desulfovibrio vulgaris* and *Methanococcus maripaludis*, but not the two monocultures, can convert lactate to methane in the absence of sulfate [27]. **(b)** The community selection process. A selection cycle starts with Newborn communities consisting of different species (different shapes), which 'mature' over time (olive arrows) to become Adult communities. During maturation, cells interact, proliferate, and possibly mutate (represented by different shades of the same shape). Adult communities with high functions (dark-purple shades) are chosen to inoculate offspring Newborns of the next selection cycle. Effective community selection requires three key elements: variation in Adult community function, differential survival based on community function, and inheritance of community function from a parent Adult to offspring Adults. **(c)** For costly community functions, inter-community selection must overcome intra-community selection. Inter-community selection occurs during community reproduction and favors communities with high community function, while intra-community selection occurs during community maturation and favors fast-growing mutants, which may contribute less to community function. Note that fast growth does not necessarily translate to high fitness in the context of community selection: although a faster grower is favored during community maturation, it may be selected against during community reproduction. **(d)** Intra-community evolution favors fast-growing (teal) mutants, which can reduce the inheritance of genotype composition between parent and offspring generations and deplete variation among offspring communities.

Intra- and inter-community selection act in opposition when community function is costly, that is, when contributing to community function reduces the growth rate of an individual (Figure 1c). Therefore, ‘cheaters’, mutants that contribute less to community function in favor of faster growth (Figure 1c, teal genotype), can evolve and spread through the population, reducing community function. Community function is also re-

### The Price equation links selection response with variation, inheritance, and selection strength

The efficacy of community selection is determined by the variation, selection, and inheritance of community function. The Price equation quantitatively links these three key elements to selection response over one cycle, the details of which are described in Box 1.

#### The Price equation describes selection response over one cycle

The Price equation [28,29] can be applied to any type of biological entity. In the following derivation, italic symbol represents an element, while bold symbol represents a vector, and  $\mathbb{E}[\dots]$  operates on a vector to obtain its expected or average value. Let  $F_i^P$  be the function (trait) value of parent  $i$  ( $1 \leq i \leq N$ ; Figure 2a).  $\mathbb{E}[F_i^O]$  is the expected function of  $i$ 's offspring **without** selection. Based on its function value  $F_i^P$ , parent  $i$  has  $W_i$  number of offspring (absolute fitness) **after** selection. Let  $\mathbb{E}[W]$  be the average number of offspring across all parents. Then, parent  $i$  has a *relative fitness*  $w_i$  ( $= W_i / \mathbb{E}[W]$ ). Let  $\mathbf{w}$  denote the vector of relative fitness for all parents. Note that the average relative fitness  $\mathbb{E}[\mathbf{w}] = 1$ .  $\delta F_i$ , the selection response of lineage  $i$  over one round of selection, is the difference between the parent function  $F_i^P$  and the offspring function after selection  $w_i \mathbb{E}[F_i^O]$ , that is,

$$\delta F_i = w_i \mathbb{E}[F_i^O] - F_i^P$$

Now we apply an algebra trick:

$$\delta F_i = \delta F_i \underbrace{- \mathbb{E}[\mathbf{w}] \mathbb{E}[F_i^O] + \mathbb{E}[F_i^O]}_{=0} = w_i \mathbb{E}[F_i^O] - \mathbb{E}[\mathbf{w}] \mathbb{E}[F_i^O] + \underbrace{\mathbb{E}[F_i^O] - F_i^P}_{\Delta F_i}$$

Note that  $\Delta F_i$  is the *transmission infidelity* of lineage  $i$ , the difference between the parent function  $F_i^P$  and its average offspring function  $\mathbb{E}[F_i^O]$  without selection (Figure 2a blue dash arrow). From the above equation, the average selection response across all parents is then given by

$$\mathbb{E}[\delta \mathbf{F}] = \mathbb{E}[\mathbf{w} \circ \{\mathbb{E}[F_1^O], \dots, \mathbb{E}[F_N^O]\}] - \underbrace{\mathbb{E}[\mathbb{E}[\mathbf{w}]]}_{=\mathbb{E}[\mathbf{w}]} \mathbb{E}[\{\mathbb{E}[F_1^O], \dots, \mathbb{E}[F_N^O]\}] + \mathbb{E}[\Delta \mathbf{F}]$$

Here,  $\mathbf{w} \circ \{\mathbb{E}[F_1^O], \dots, \mathbb{E}[F_N^O]\}$  is the vector of the element-wise product of the two vectors, and  $\mathbb{E}[\Delta \mathbf{F}]$  is the average transmission infidelity across all lineages. Recall that for two vectors  $\mathbf{x}$  and  $\mathbf{y}$ , their covariance  $\text{Cov}(\mathbf{x}, \mathbf{y}) = \mathbb{E}[\mathbf{x} \circ \mathbf{y}] - \mathbb{E}[\mathbf{x}] \mathbb{E}[\mathbf{y}]$ . Thus, the above equation becomes

$$\mathbb{E}[\delta \mathbf{F}] = \text{Cov}(\mathbf{w}, \{\mathbb{E}[F_1^O], \dots, \mathbb{E}[F_N^O]\}) + \mathbb{E}[\Delta \mathbf{F}]$$

where  $\text{Cov}(\mathbf{w}, \{\mathbb{E}[F_i^O]\})$  describes the covariance between parent fitness with the respective average offspring function. This covariance term can be transformed to covariance between parent fitness and parent function if we assume a linear relationship between parent and offspring function (Figure 2b)

$$\mathbb{E}[F_i^O] = h^2 F_i^P + \theta_i + \varepsilon_i$$

where  $h^2$  is the slope of linear least squares regression between parent trait and average offspring trait,  $\theta_i$  is the intercept, and  $\varepsilon_i$  is the residual.  $h^2$ , also known as the ‘biometric heritability’, attempts to measure the proportion of  $F$ 's variance that is heritable [30,31]. With a linear parent-offspring relation, fitness does not covary with residual (i.e.  $\text{Cov}[\mathbf{w}, \boldsymbol{\varepsilon}] = 0$ ), and the Price equation becomes

$$\mathbb{E}[\delta \mathbf{F}] = h^2 \text{Cov}(\mathbf{w}, \mathbf{F}^P) + \mathbb{E}[\Delta \mathbf{F}] \quad \text{where } h^2 = \frac{\text{Cov}(\mathbf{F}^P, \{\mathbb{E}[F_i^O]\})}{\text{Var}(\mathbf{F}^P)}$$

In summary, the Price equation (Figure 2c) describes the selection response of any entity over one cycle ( $\mathbb{E}[\delta \mathbf{F}]$ ) as a function of selection strength ( $\text{Cov}(\mathbf{w}, \mathbf{F}^P)$ ), and inheritance in the form of both biometric heritability ( $h^2$ ) and transmission infidelity ( $\mathbb{E}[\Delta \mathbf{F}]$ ). Not surprisingly, biometric heritability is linked to transmission infidelity  $\Delta \mathbf{F}$ . Since  $\{\mathbb{E}[F_i^O]\} = \mathbf{F}^P + \Delta \mathbf{F}$ ,

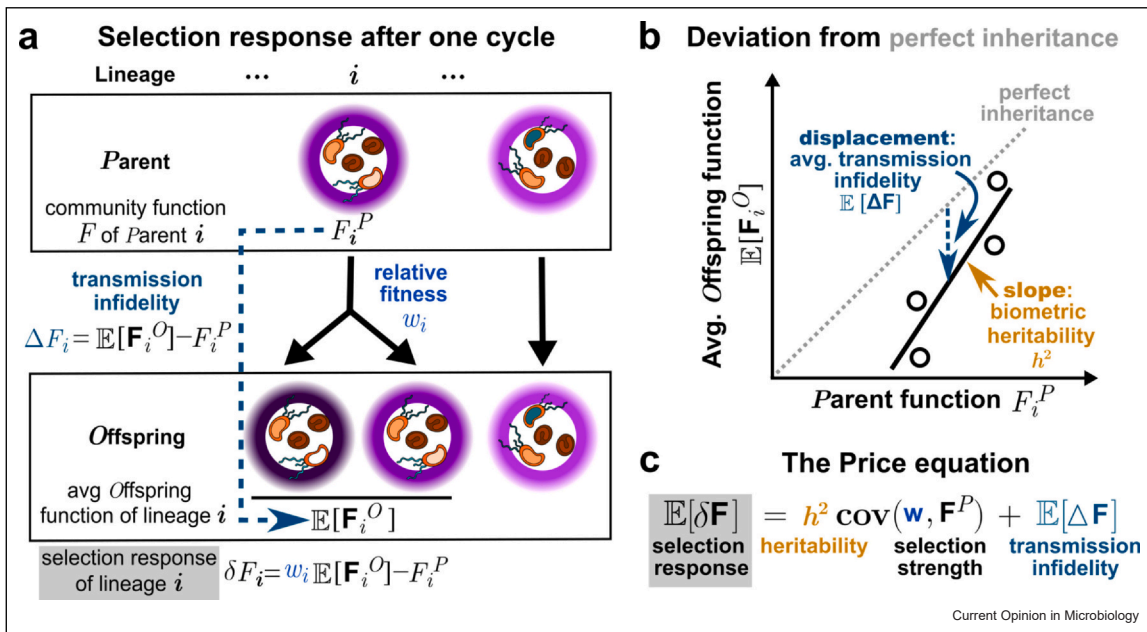
$$h^2 = \frac{\text{Cov}(\mathbf{F}^P, \mathbf{F}^P + \Delta \mathbf{F})}{\text{Var}(\mathbf{F}^P)} = \frac{\text{Var}(\mathbf{F}^P) + \text{Cov}(\mathbf{F}^P, \Delta \mathbf{F})}{\text{Var}(\mathbf{F}^P)} = 1 + \frac{\text{Cov}(\mathbf{F}^P, \Delta \mathbf{F})}{\text{Var}(\mathbf{F}^P)}$$

duced if a fast-growing species outcompetes a slow-growing species essential for community function. A telltale of a costly community function is that it will decline in the absence of community selection due to intra-community selection. Costly community functions will be the focus of this review since they are often observed in experimental systems (e.g. [11,16]), and are more difficult to improve compared with noncostly community functions.

When applied to community selection, the Price equation (Figure 2c, Box 1) describes how the average community function changes after one selection cycle ([28,29], see critiques in Ref. [32]):

$$\underbrace{\mathbb{E}[\delta \mathbf{F}]}_{\text{Selection response after one selection cycle}} = \underbrace{h^2}_{\text{Biometric heritability}} \times \underbrace{\frac{\text{Cov}(\mathbf{w}, \mathbf{F}^P)}{\text{Var}(\mathbf{F}^P)}}_{\text{Selection strength}} + \underbrace{\mathbb{E}[\Delta \mathbf{F}]}_{\text{Avg. transmission infidelity}} \quad (1)$$

Figure 2



The Price equation links selection response to key elements of evolution. **(a)** Mathematical nomenclature for selection response of a community function after one round of selection. The higher the community function, the darker the shade of purple ring. Lineage  $i$  is used as an example. **(b)** The inheritance of community function can be described by biometric heritability  $h^2$  (the slope of parent–offspring linear regression, gold) and average transmission infidelity  $\mathbb{E}[\Delta F]$  (the average displacement of the regression line from perfect inheritance, teal). Perfect inheritance is achieved when parent and offspring functions are identical (dotted line, slope  $h^2 = 1$  and displacement  $\mathbb{E}[\Delta F] = 0$ ). **(c)** The Price equation.

Here,  $\mathbb{E}[\delta F]$  is the selection response after one cycle — the difference between the expected function of all parents before selection and the expected function of all offspring after selection (Figure 2a, gray arrow).  $h^2$  is the biometric heritability of community function, the slope of the linear least squares regression between parent function and average offspring function (Figure 2b, gold).  $\text{Cov}(\mathbf{w}, \mathbf{F}^P)$  is the covariance between parent relative fitness  $\mathbf{w}$  (the number of offspring from a parent divided by the average number of offspring after selection) and parent function  $\mathbf{F}^P$ .  $\mathbb{E}[\Delta F]$  is the average transmission infidelity (teal dash arrow in Figure 2b), the average change in community function from a parent to its offspring without inter-community selection. Transmission infidelity is driven by eco-evolutionary dynamics and stochastic factors, which alter genotype and species compositions during community maturation and reproduction. For example, as cheaters rise in frequency, community function will tend to decline from parent to offspring, leading to a negative average transmission infidelity  $\mathbb{E}[\Delta F]$  (Figure 2b, teal arrow).

The Price equation contains all three key elements of evolution. Selection strength is described by  $\text{Cov}(\mathbf{w}, \mathbf{F}^P)$ , the covariance of parent fitness with parent community function. Inheritance of community function (Figure 2b) is described by biometric heritability  $h^2$  and the average transmission infidelity  $\mathbb{E}[\Delta F]$ , which are

interconnected (Box 1, last paragraph). Variation in parent community function  $\text{Var}(\mathbf{F}^P)$  is more subtle, but is included in biometric heritability  $h^2$  (Box 1, last equation), as well as in selection strength since covariance between fitness and function only makes sense in the presence of variation in function. Over the course of selection, as the fittest communities are selected, variation among communities decreases, which slows down the improvement of community function. Interestingly, biometric heritability also becomes low, possibly further reducing improvement [10,11,33].

### Understanding community selection: community function determinants and landscape

Similar to phenotypes of individuals, which are shaped by genotype and environment, community function is shaped by ‘community function determinants’ (‘determinants’). Determinants are factors that vary among communities, and whose variation causes community function to vary [34]. Examples of determinants include genotype compositions within each species (‘genotype determinant’), species composition of a community (‘species composition determinant’), and environmental variables (‘environmental determinant’). Measurement noise is also a determinant since measured community function is the actual substrate of selection. Some determinants, such as environmental variability and



measurement noise, are not heritable from parent to offspring generations. Some determinants, such as the genotype composition of each species, are heritable to a certain degree under some conditions. Other determinants, such as species composition, may or may not be heritable, as we will discuss below. Overall, understanding the variation and inheritance of determinants allows us to understand the variation and inheritance of community function. Below, we will discuss genotype and species composition determinants in greater detail, as they are likely important for most community functions.

#### Genotype composition determinants are often heritable

The genotype determinant of a species in community can be calculated at the Newborn stage as the average genotype value, that is, the average phenotype, of individuals within the species [34]. Although genotype values, such as growth rates, should not be averaged in general, averaging turns out to be valid when intra-community evolution is slow [34]. A genotype determinant, such as the average fraction of resource invested on community function, can be inherited from a parent Newborn to its offspring Newborns: For example, a cheater-dominated parent community will likely generate cheater-dominated offspring communities [34]. However, inheritance can be compromised by intra-community evolution during community maturation, and by stochastic sampling of genotypes during community reproduction.

#### Species composition determinants may or may not be heritable depending on ecological dynamics

Species composition determinants can be defined as the fractional abundance of species in a Newborn if the total biomass of the Newborn is held constant. Species composition determinant is often defined in a Newborn rather than in an Adult for two reasons: first, Newborn species composition can influence community function as initial conditions for ordinary differential equations. For example, the total amount of product accumulated in an Adult can depend on Newborn species composition [34]. Second, Adult species composition may not serve as a determinant if different Newborn species compositions become identical during community maturation. This can arise due to an ‘ecological attractor’, a stable equilibrium such that after small perturbations, the system will return to that equilibrium (Figure 3a i). Note that an equilibrium is not always an attractor, for example, with a semistable attractor, a small perturbation may or may not drive the system to a new equilibrium (Figure 3a iii, dotted line). If the community has only one attractor and if the attractor is reached during maturation, then species composition is non-heritable (Figure 3a i). If the community has multiple attractors, species composition becomes a heritable determinant (Figure 3a ii). In communities that lack stable attractors, species composition may or may not be heritable. For example, species composition is heritable in communities with a semistable attractor (Figure 3a iii), but

nonheritable in a community exhibiting chaotic dynamics (Figure 3a iv).

#### Understanding heritability of community function in terms of heritability of determinants

Community function can be approximated as the linear sum of its determinants, if intra-community evolution is sufficiently slow and if determinants are narrowly distributed among communities [48]. Consider a simple case where community function  $F$  has two determinants:  $x$  that is heritable (positive  $h_x^2$ , the slope of parent–offspring regression of  $x$ ) and  $y$  that is nonheritable ( $h_y^2 = 0$ ). If the linear approximation holds, then for community  $i$ , its function is

$$F_i \approx \alpha x_i + \beta y_i + \theta + \varepsilon_i \quad (2)$$

where  $\alpha$  and  $\beta$  express how much the corresponding determinant affects community function,  $\theta$  is the intercept, and  $\varepsilon$  is the residual of linear regression. If, additionally, the residual term is uncorrelated with determinants and if the determinants are independent of each other, it can be shown that the heritability of community function  $F$  is [48]

$$h^2 \approx \frac{\alpha^2 h_x^2 \text{Var}(x)}{\text{Var}(F)}. \quad (3)$$

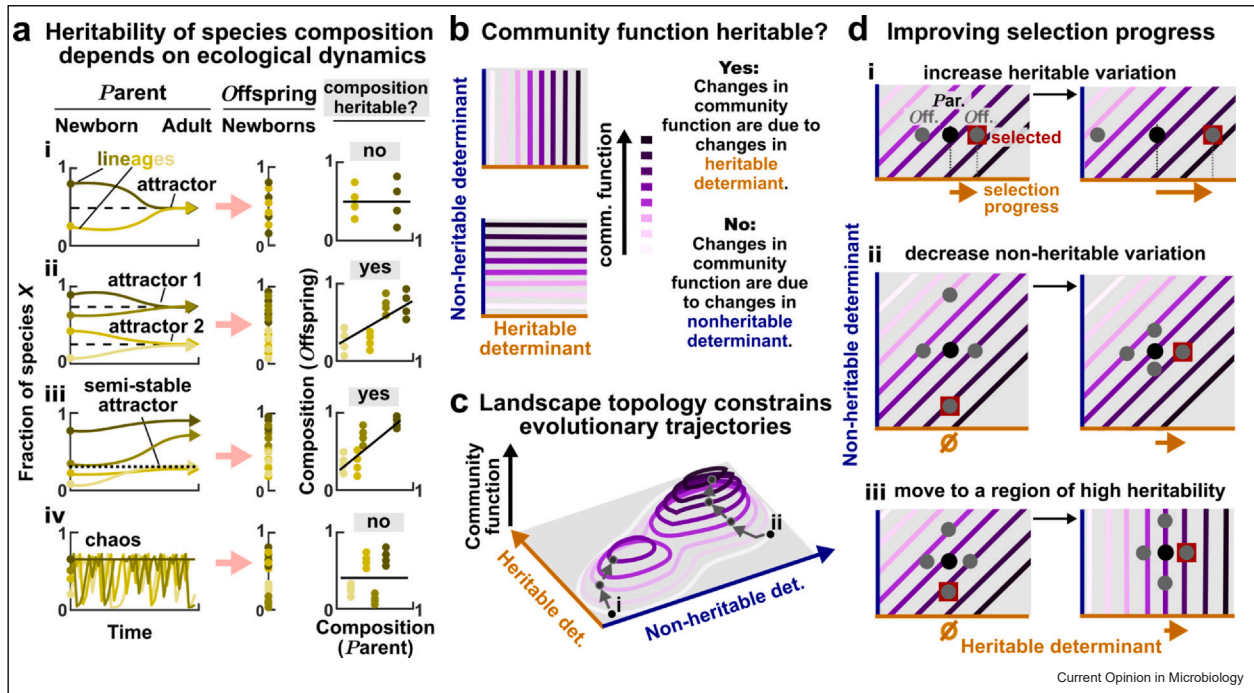
This is intuitive, as it states that the heritability of a community function is determined by variation in the heritable determinant relative to variation in the community function. This can then be substituted into Price equation (1) to estimate selection response.

#### Constructing the community function landscape from determinants

We can visualize how community function varies with determinants, heritable and nonheritable, using a community function landscape (Figure 3b and c). For any community on the landscape, its determinants can be read off the axis, while its function can be read from community function contours (purple lines, varying shades corresponding to different levels of community function). If communities to be selected are scattered in a region where contours are parallel to the axis of the heritable determinant (Figure 3b, bottom), then variations in community function are solely due to variations in the nonheritable determinant, and thus selection will not be effective. Conversely, in regions where contours are perpendicular to the axis of the heritable determinant (Figure 3b, top), selection will be effective.

The effectiveness of community selection can be visualized as *selection progress*, change in the heritable determinant over one selection cycle, which defines the heritable portion of the selection response in the Price equation. Note that although selection response is

Figure 3



Understanding selection progress in terms of community function determinants and landscape. **(a)** Heritability of the species composition determinant depends on ecological dynamics. Each shade represents an independent community lineage. **(i)** Species composition varies among parent Newborns, but during maturation moves toward one and only one attractor (dashed line in the left panel). Parent Adults are reproduced through pipetting (pink arrow), which introduces stochastic variation into the offspring Newborns (center panel). In this case, species compositions in offspring Newborns are uncorrelated with those in parent Newborns, and thus species composition is not heritable (right panel). **(ii)** With multiple attractors, a parent Newborn will tend toward the closest attractor, and species composition is heritable. **(iii)** Although communities with a composition slightly above or below the semistable attractor (dotted line) will reach different equilibria, species composition is heritable overall. **(iv)** A chaotic system where species composition is nonheritable. The dynamics is based on a two-species community where one species has a constant population (set at 0.5) and the other grows according to the logistic difference equation:  $x_{t+1} = 4x_t(1 - x_t)$ . The initial  $x$  was set to 0.1, 0.3, 0.48, and 0.75, and maturation time = 40. The system is chaotic from most initial conditions, but when the parent Newborn fraction = 0.75, species composition remains constant. Systems without attractors can exhibit other behaviors, such as remaining at the present value indefinitely until perturbed, and in this case, species composition determinant will be heritable. **(b)** Community function heritability is determined by the orientation of community function contours on the community function landscape with respect to the axes of heritable and nonheritable determinants. **(c)** Landscape topology, including the presence of attractors, constrains evolutionary trajectories. This makes selection outcomes dependent on starting positions. Community function landscapes can contain multiple 'peaks' of high community function and 'valleys' of low community function between peaks. Thus, depending on their starting points, communities can get stuck at lower fitness peaks (i), or climb to the global maximum (ii). **(d)** Strategies for improving selection progress. In this graph, a community's determinants, defined in the Newborn stage, can be read off the axes, while its function, defined in the Adult stage, can be read off the community function contours. Selection progress can be measured as a change in any heritable determinant over one selection cycle (represented by the length of an orange arrow, or  $\emptyset$  if no progress). Selection efficacy can be improved by increasing variation in the heritable determinant (i), or decreasing variation in the nonheritable determinant (ii). (iii) Selection progress also increases by moving communities from regions of low inheritance to regions of high inheritance.

described by the Price equation, only the heritable portion — selection progress — can be propagated into future cycles. Thus, selection becomes effective when variation in heritable determinants ('heritable variation') is increased (Figure 3d i), when variation in nonheritable determinants ('nonheritable variation') is decreased (Figure 3d ii), or when communities are moved from a low-inheritance region to a high-inheritance region [34] (Figure 3d iii). Even when a landscape cannot be

visualized, whenever selection is deemed ineffective, we can test different strategies and choose the strategy conferring the highest biometric heritability [34]. This method worked effectively in Xie et al.'s (2021) study, as did methods that reduce nonheritable variations, such as reproducing the parent community via a cell sorter to sort precise biomass of each species into Newborns, instead of pipetting which introduces fluctuations in species composition [35].

**Table 1****Manipulation of an experimental variable often exerts opposite effects on key elements for selection success.**

Experimental design and manipulations	Pros	Cons
(a) Increase the total number of communities under selection	<ul style="list-style-type: none"> <li>Increases variation without sacrificing inheritance</li> </ul>	<ul style="list-style-type: none"> <li>Experimentally more challenging</li> <li>May not be very effective [36–38] (e.g. scaling with <math>\sqrt{\log(n)}</math> where <math>n</math> is the number of communities [36])</li> </ul>
(b) Reduce population size in Newborns	<ul style="list-style-type: none"> <li>Increases inheritance by reducing intra-community evolution (Figure 1d)</li> <li>Increases stochastic variations in species and genotype compositions among offspring</li> </ul>	<ul style="list-style-type: none"> <li>Decreases variation by restricting the number of new mutants arising</li> <li>Decreases inheritance by increasing stochastic variations among Newborns (e.g. rare species or rare genotypes can be lost by chance)</li> </ul>
(c) Reduce mutation rate or resource availability	<ul style="list-style-type: none"> <li>Increases variation and inheritance by reducing intra-community evolution</li> </ul>	<ul style="list-style-type: none"> <li>Decreases variation by reducing mutational opportunities</li> </ul>
(d) Increase maturation time	<ul style="list-style-type: none"> <li>Increases inheritance by allowing species composition to stabilize</li> </ul>	<ul style="list-style-type: none"> <li>Decreases variation and inheritance by increasing intra-community evolution (Figure 1d)</li> </ul>
(e) Promote species coexistence (e.g. engineering dependency between species)	<ul style="list-style-type: none"> <li>Increases inheritance (e.g. by preventing species loss)</li> </ul>	<ul style="list-style-type: none"> <li>Can trap communities in low-inheritance regions of community function landscape [34]</li> </ul>
(f) Selectively add or remove species during community reproduction	<ul style="list-style-type: none"> <li>Increases variation by creating new community compositions [39]</li> </ul>	<ul style="list-style-type: none"> <li>Decreases inheritance between parent offspring generations</li> </ul>
(g) Choose fewer Adults to reproduce	<ul style="list-style-type: none"> <li>Increases selection strength</li> </ul>	<ul style="list-style-type: none"> <li>Decreases variation by omitting communities that have desired genotypes but display low function driven by for example, unfavorable measurement noise [35]</li> </ul>
<p>Selection outcome can become highly variable among replicate runs if selection is too stringent, although community function improvement becomes more limited if selection is too lenient [40]. Simulations suggest that a large range of intermediate-strength selection regimes produce similar selection efficacy [35].</p>		
(h) 'Migrant pool' reproduction (mixing parent Adult communities before reproduction)	<ul style="list-style-type: none"> <li>Can potentially pull together complementary members from different parents</li> </ul>	<ul style="list-style-type: none"> <li>Decreases variation by homogenizing parents (e.g. allowing cheaters to spread)</li> <li>Decreases inheritance by making offspring resemble the migrant pool rather than parent</li> </ul>
Versus 'Propagule' reproduction (no mixing of parent Adult communities before reproduction)	<ul style="list-style-type: none"> <li>Increases inheritance from parent to offspring</li> <li>Maintains some variation</li> </ul>	<ul style="list-style-type: none"> <li>The initial ecological variations among communities will be rapidly lost as the top-functioning lineages take over. Unless sufficient variation is introduced through mutation or migration, selection response rapidly plateaus [11,41,42]</li> </ul>

### Challenges and solutions for designing effective selection strategies

Community selection is challenging because manipulating an experimental variable often exerts opposing effects. For example, increasing variation often leads to reduced inheritance (Table 1, f), and increasing selection strength leads to reduced variation (Table 1, g). Intra-community evolution also exerts opposing effects: although it supplies new mutations critical for improving community function, it also favors fast-growing cheater mutants. As cheaters take over, all communities will be similarly dominated by cheaters (low variation among communities), and community function will decline

from parent to offspring (low inheritance from parent to offspring) (Figure 1d). Indeed, when selecting for chitin-degrading communities, as chitin degradation became faster, maturation time had to be shortened to prevent nondegrading species from taking over communities [16]. Therefore, when we try to increase heritable variation (Figure 3d i), intra-community evolution should not be significantly accelerated. Thus, strategies that tend to increase the rate of intra-community evolution, such as increasing mutation rate, increasing Newborn population size, or increasing the number of doublings by supplementing extra resources or prolonging maturation time, may not be effective (Table 1). In

**Table 1**, we summarize a (nonexhaustive) list of experimental variables that can be manipulated, and the pros and cons associated with each.

How might we implement the above principles to improve community function? Owing to the commonly occurring trade-off between variation and inheritance (**Table 1**), multiple manipulations are sometimes needed to promote both variation and inheritance, as demonstrated by previous theoretical studies. For example, Chang et al. [39] first ensured inheritance by allowing communities to reach an equilibrium, so that parent and offspring communities would share similar species compositions. Then, heritable variations were introduced to a subset of offspring communities by, for example, introducing or removing species. As another example, Vessman et al. [38] increased heritable variation by introducing species into or removing species from some of the Newborns, and to facilitate inheritance, species composition was adjusted to a target value in each round of selection. Both methods outperformed the standard community reproduction methods, that is, migrant pool or propagule reproduction (**Table 1**). In general, if communities harbor or evolve [43] attractors (e.g. [44–46]), evolutionary trajectories may be restricted along the attractors [34], and exploration of the wider community function landscape is prevented. In this case, periodically perturbing and destabilizing species composition, for example, by inducing chaos [47] and then allowing composition to stabilize, could allow communities to move between different heritable species compositions. This could move communities out of local optima (**Figure 3c**) or move communities from regions of low heritability to regions of high heritability (**Figure 3d iii**, [34]).

In conclusion, many experimental variables in community selection studies can have concurrent positive and negative effects on variation and inheritance, which makes improving community function difficult. Even increasing variation or inheritance does not always improve community function. For example, while heritable variation can drive selection progress, nonheritable variation cannot. Likewise, although inheritance is important for securing selection progress, periodically reducing inheritance by introducing heritable variation is important for improving community function.

The field of community selection is still in its infancy. Owing to high computational demands, theoretical work on selecting ecologically complex communities generally has not considered evolution (e.g. [39,41]), or has considered evolution in an unrealistic manner (e.g. [36]). Other theoretical work considered ecological–evolutionary dynamics, but was restricted to simple two-species communities (e.g. [34,35,37,43]). An important next step for

theoretical studies is to integrate evolution and ecology in many-species communities. This will facilitate a better understanding of how community function evolves in complex systems and the design of more effective selection strategies. Empirical work on complex communities has been largely limited, with any improvement in community function likely relying on selection of preexisting standing variation across subcommunities. This causes any improvement to level off quickly as variation is depleted by selection (e.g. [11,41,42]). Thus, an important next step for experimental studies is to systematically test different strategies, such as those listed in **Table 1**, for their ability to improve community function. More specifically, we need to understand the effect of each strategy on variation and inheritance, as this will allow us to identify strategies that improve one without sacrificing the other, or improve both simultaneously. Together, these efforts will lay the foundation for a mature discipline of microbial biotechnology, with applications in domains from waste treatment to medicine.

## Funding

This work was supported by a professorship from Academy of Medical Sciences (AMS) and the Government Department of Business, Energy and Industrial Strategy (BEIS), and a fellowship from Wolfson Foundation and Royal Society.

## Data Availability

Data will be made available on request.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

We thank Shahriar Shahriar and members of the Shou lab, especially Alex Yuan, Roman Zapien-Campos, Andre Ribeiro, Cansu Oktem, and Patricia Gimeno Le Paih, for comments. AY and RZC additionally helped with document formatting and mathematical nomenclature, respectively.

## References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

•• of outstanding interest

1. Ghoul M, Mitri S: **The ecology and evolution of microbial competition**. *Trends Microbiol* 2016, **24**:833–845.
2. Nadell CD, Drescher K, Foster KR: **Spatial structure, cooperation and competition in biofilms**. *Nat Rev Microbiol* 2016, **14**:589–600.
3. West SA, et al.: **Social evolution theory for microorganisms**. *Nat Rev Microbiol* 2006, **4**:597–607.



4. Zhou K, *et al.*: **Distributing a metabolic pathway among a microbial consortium enhances production of natural products.** *Nat Biotechnol* 2015, **33**:377-383.
5. Jia N, *et al.*: **Comparative genomics and metabolomics analyses of the adaptation mechanism in *Ketogulonigenium vulgare-Bacillus thuringiensis* consortium.** *Sci Rep* 2017, **7**:46759.
6. Lawley TD, *et al.*: **Targeted restoration of the intestinal microbiota with a simple, defined bacteriotherapy resolves relapsing *Clostridium difficile* disease in mice.** *PLoS Pathog* 2012, **8**:e1002995.
7. Kato S, *et al.*: **Effective cellulose degradation by a mixed-culture system composed of a cellulolytic *Clostridium* and aerobic non-cellulolytic bacteria.** *FEMS Microbiol Ecol* 2004, **51**:133-142.
8. Piccardi P, Vessman B, Mitri S: **Toxicity drives facilitation between 4 bacterial species.** *Proc Natl Acad Sci* 2019, **116**:15979-15984.
9. Lindemann SR, *et al.*: **Engineering microbial consortia for controllable outputs.** *ISME J* 2016, **10**:2077-2084.
10. Arora J, Brisbin MAM, Mikheyev AS: **Effects of microbial evolution dominate those of experimental host-mediated indirect selection.** *PeerJ* 2020, **8**:e9350.
11. Chang C-Y, *et al.*: **Artificially selecting bacterial communities using propagule strategies.** *Evolution* 2020, **74**:2392-2403.  
This experimental study demonstrated that intra-community evolution and diminishing variation caused community selection to fail in two systems.
12. Goodnight CJ: **Experimental studies of community evolution I: the response to selection at the community level.** *Evolution* 1990, **44**:1614-1624.
13. Goodnight CJ: **Experimental studies of community evolution II: the ecological basis of the response to community selection.** *Evolution* 1990, **44**:1625-1636.
14. Mueller UG, *et al.*: **Artificial selection on microbiomes to breed microbiomes that confer salt tolerance to plants.** *Msystems* 2021, **6**:e01125-e01221.
15. Panke-Buisse K, *et al.*: **Selection on soil microbiomes reveals reproducible impacts on plant function.** *ISME J* 2015, **9**:980-989.
16. Wright RJ, Gibson MI, Christie-Oleza JA: **Understanding microbial community dynamics to improve optimal microbiome selection.** *Microbiome* 2019, **7**:1-14.  
In this work, the authors troubleshooted the low efficacy of community selection. They found that the maturation time needed to be progressively shortened to prevent cheaters from taking over the community.
17. Raynaud T, *et al.*: **Effect of the reproduction method in an artificial selection experiment at the community level.** *Front Ecol Evol* 2019, **7**:416.
18. Swenson W, Arendt J, Wilson DS: **Artificial selection of microbial ecosystems for 3-chloroaniline biodegradation.** *Environ Microbiol* 2000, **2**:564-571.
19. Swenson W, Wilson DS, Elias R: **Artificial ecosystem selection.** *Proc Natl Acad Sci* 2000, **97**:9110-9114.
20. Jochum MD, *et al.*: **Host-mediated microbiome engineering (HMME) of drought tolerance in the wheat rhizosphere.** *PLoS One* 2019, **14**:e0225933.
21. Mueller UG, Sachs JL: **Engineering microbiomes to improve plant and animal health.** *Trends Microbiol* 2015, **23**:606-617.
22. Arias-Sanchez FI, Vessman B, Mitri S: **Artificially selecting microbial communities: if we can breed dogs, why not microbiomes?** *PLoS Biol* 2019, **17**:e3000356.
23. Goodnight CJ: **Evolution in metacommunities.** *Philos Trans R Soc B: Biol Sci* 2011, **366**:1401-1409.
24. Sánchez A, *et al.*: **Directed evolution of microbial communities.** *Annu Rev Biophys* 2021, **50**:323-341.
25. Shou W: **Acknowledging selection at sub-organismal levels resolves controversy on pro-cooperation mechanisms.** *Elife* 2015, **4**:e10106.
26. Lewontin RC: **The units of selection.** *Annu Rev Ecol Syst* 1970, **1**:1-18.
27. Hillesland KL, Stahl DA: **Rapid evolution of stability and productivity at the origin of a microbial mutualism.** *Proc Natl Acad Sci* 2010, **107**:2124-2129.
28. Price GR: **Selection and covariance.** *Nature* 1970, **227**:520-521.
29. Price GR: **Extension of covariance selection mathematics.** *Ann Hum Genet* 1972, **35**:485-490.
30. Bourrat P: **Unifying heritability in evolutionary theory.** *Stud Hist Philos Sci* 2022, **91**:201-210.
31. Jacquard A: **Heritability: one word, three concepts.** *Biometrics* 1983, **39**:465.
32. Van Veelen M, *et al.*: **Group selection and inclusive fitness are not equivalent; the Price equation vs. models and statistics.** *J Theor Biol* 2012, **299**:64-80.
33. Blouin M, *et al.*: **Levels and limits in artificial selection of communities.** *Ecol Lett* 2015, **18**:1040-1048.
34. Xie L, Shou W: **Steering ecological-evolutionary dynamics to improve artificial selection of microbial communities.** *Nat Commun* 2021, **12**:6799.  
This paper described the concept of community function determinants and landscape, as well as how to improve selection efficacy by shifting communities to a region of higher inheritance.
35. Xie L, Yuan AE, Shou W: **Simulations reveal challenges to artificial community selection and possible strategies for success.** *PLoS Biol* 2019, **17**:e3000295.
36. Fraboul J, Biroli G, De Monte S: **Artificial selection of communities drives the emergence of structured interactions.** *J Theor Biol* 2023, **571**:111557.
37. Lee J, Shou W, Park HJ: **Condition for successful artificial selection on microbial collectives and optimal experimental setup.** *bioRxiv* 2023,531234.
38. Vessman B, *et al.*: **Novel artificial selection method improves function of simulated microbial communities.** *bioRxiv* 2023,523165.
39. Chang C-Y, *et al.*: **Engineering complex communities by directed evolution.** *Nat Ecol Evol* 2021, **5**:1011-1023.  
By performing *in-silico* selection of microbial metacommunities, this work identified a strategy to increase selection efficacy: iterations of allowing communities to reach compositional equilibrium to increase inheritance, followed by applying diverse perturbations to introduce variation.
40. Zaccaria M, *et al.*: **Partner-assisted artificial selection of a secondary function for efficient bioremediation.** *iScience* (9) 2023, **26**.
41. Penn A, Harvey I: **The Role of Non-genetic Change in the Heritability, Variation, and Response to Selection of Artificially Selected Ecosystems.** MIT Press; 2004:352.
42. Williams HTP, Lenton TM: **Artificial selection of simulated microbial ecosystems.** *Proc Natl Acad Sci* 2007, **104**:8919-8923.
43. Douclier G, *et al.*: **Eco-evolutionary dynamics of nested Darwinian populations and the emergence of community-level heredity.** *Elife* 2020, **9**:e53433.  
Modelling communities of competing microbes showed that selecting community-level traits can drive the evolution of an attractor, allowing competing microorganisms to stably co-exist rather than out-competing each other.
44. Gonze D, *et al.*: **Microbial communities as dynamical systems.** *Curr Opin Microbiol* 2018, **44**:41-49.
45. Wilson DS: **Complex interactions in metacommunities, with implications for biodiversity and higher levels of selection.** *Ecology* 1992, **73**:1984-2000.
46. Wright ES, Gupta R, Vetsigian KH: **Multi-stable bacterial communities exhibit extreme sensitivity to initial conditions.** *FEMS Microbiol Ecol* 2021, **97**:fiab073.
47. Karkaria BD, *et al.*: **Chaos in synthetic microbial communities.** *PLoS Comput Biol* 2022, **18**:e1010548.

48. Xie L, Yuan AE, Shou W: **A quantitative genetics framework for understanding the selection response of microbial communities.** *bioRxiv* 2023,2023-2110.

## Glossary

**Entity:** A biological entity ('entity') is a biological structure with a boundary, such that the birth, growth, survival, or death of an entity is separable from that of other similar entities. This is due to chemical or physical coupling within an entity and the lack of equivalent coupling between entities.

**Interaction:** Instances where one community member affects the physiology of another, for example, by releasing nutrients, toxins, or signaling molecules.

**Community function:** Community-level activities that are not the sum of activities in monoculture, arising from interactions between community members. Community function is generally defined in the Adult stage.

**Costly community function:** A community function is costly if contributing to this function slows down cell growth in at least one member species.

**Community function determinants (determinants):** Factors whose variation drives variation in community function. Examples include genotype composition, species composition, and environmental conditions. Determinants are generally defined in the Newborn stage, except for factors that are not present at initiation, e.g., measurement noise of community function. The inheritance of determinants is on a spectrum, i.e., determinants can be heritable or non-heritable to greater or lesser extents.

**Community selection:** The process of breeding new communities by artificially selecting for high community function.

**Newborn and Adult community:** A Newborn community ("Newborn") is the initial low cell density community at the beginning of a selection cycle. A Newborn grows and matures into an Adult community ("Adult"). A chosen Adult community is used to inoculate Newborn communities of the next selection cycle.

**Parent and offspring community:** A parent community is a community in the parent generation that is selected for reproduction. The community is then partitioned into multiple offspring communities. Both parent and offspring communities have Newborn and Adult stages. To measure selection response, we compare community functions of parent Adult and offspring Adult. To measure selection progress, we compare heritable determinants of parent Newborn and offspring Newborn.

**Evolution and selection:** Selection is the differential survival of entities, and is part of evolution. Evolution also includes the emergence of new traits through, for example, mutations.

**Intra- and Inter-community selection:** Intra-community selection is natural selection acting on individuals within a community during maturation, and favors faster growers. Inter-community selection is artificial selection acting on communities during community reproduction, and favors high community function.

**Inheritance:** Resemblance between parent and offspring.

**Biometric Heritability:** The slope of a linear regression between parent and average offspring trait values.

**Transmission Infidelity:** The difference between the parent trait and the average offspring trait in the absence of selection.

**Selection Strength:** The covariance between the number of offspring and parent trait, i.e., how many offspring an Adult parent produces based on its trait.

**Ecological attractor:** The species composition that Newborns with different initial compositions converge on during community maturation.